



303 –LANGUAGE TEXT IDENTIFICATION

TEAM INFORMATION

Team Name: _____

Results Email: _____

Examination Time Frame: _____ to _____

INSTRUCTIONS

Description: Examiners must develop and document a methodology used to identify the language presented in the documents in the **303_Language_Text_Identification_Challenge2008**.

Report the detailed explanation of your process (software or technique) used to determine the language and document information.

Total Weighted Points: 60 Total Points available per entry – Total 300 Points Available

1. **Answers** – Fill in the chart below with your findings. *As a Forensic Challenge, consider that your answers will have to have enough detail for the Findings and Methodology of your examination to satisfy questioning in a court of law.*
2. **Methodology** – Provide a meticulously detailed explanation of your process. Be sure to include a step action that our reviewers can follow to reproduce your work for authenticity including tools and techniques.

INTERNAL REVIEWER USE ONLY

Reviewer:

Points Awarded:

Date:

Review Period: _____ to _____

Completed: ☐ Yes

☐ No

☐ Partial **Team Exaflop 303**

Page 1 of 4 11/6/2008

REPORT OF EXAMINATION

<Answer Area>

File Name

Original Language

1.

2.

3.

4.

5.

6.

7.

Analysis that validates your findings:

Original Language Identification Method

1.

2.

3.

4.

5.

6.

7.

Please attach additional sheets as needed.

Page _____ of _____ Initials _____

METHODOLOGY / NOTES FORM

Tool Information

Type	Name	Publisher
------	------	-----------

Type	Name	Publisher
------	------	-----------

Type	Name	Publisher
------	------	-----------

Commercial / Open Source

Site:

Type	Name	Publisher
------	------	-----------

Type	Name	Publisher
------	------	-----------

Type	Name	Publisher
------	------	-----------

Commercial / Open Source

Site: _____

Type	Name	Publisher
------	------	-----------

Type	Name	Publisher
------	------	-----------

Type	Name	Publisher
------	------	-----------

Commercial / Open Source

Site: _____

[illegible]

As per the challenge question I did not include any of the translations. Below is my mythology for determining which language was presented in each document.

- 1) Use Microsoft Word 2007 to open the files.
- 2) Examine each document and copy the text into <http://babelfish.yahoo.com>
- 3) Used my personal knowledge of how particular languages are written and selected a translation package
- 4) Reviewed the results – IF the result was readable in English Proceed to step 5 IF NOT go back to step 3.
- 5) Test the same copied text in <http://translate.google.com> (Google has an option to select the input language as “Dialect Language” which helps automate the language determination process)
- 6) Is the translation from step 2 similar to step 5? IF YES move to step 7, IF NO move to step 1. (In comparing the different free translation engines – I’ve found that most are coded differently and some reveal more info than others)
- 7) Test the same copied text into Microsoft Word 2007 and use the internal translation technique (Review → Translate) with the same parameters used in steps 2 and 5. IF the translation is similar to steps 2 and 5 move to step 8. IF NO determine if the changes are acceptable – IF NO then go to step 2.
- 8) Answer found.

Text Identification:

Document1Tx.doc – Chinese

Document2Tx.doc - Russian

Document3Tx.doc – Korean

Document4Tx.doc – Arabic

Document5Tx.doc – Japanese